DOCUMENT RESUME

ED 185 111 **TM 800 124**

AUTHOR Garcia-Quintana, Roan A.; Ryan, Joseph P.

TITLE What Does It Mean for Items to Fit a Latent Trait Psychometric Model: A Comparison of Model Fitting with Content Validation.

PUB DATE Feb 79
NOTE 14p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Kiawah Island, SC, February 20-24, 1979).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Course Objectives; *Criterion Referenced Tests; Elementary School Mathematics; Factor Analysis; *Goodness of Fit; Grade 6; Intermediate Grades; *Item Analysis; Test Interpretation; *Test Items; *Test Validity
IDENTIFIERS *Content Validity; Mastery (Science Research Associates); *Rasch Model; South Carolina

ABSTRACT

A sixth-grade mathematics test was evaluated in two different ways, and the evaluations were compared. One method of evaluation was to use the Rasch model for item analysis and to accept or reject items on the basis of the extent to which the data fit the model. The other method of evaluation was to ask school superintendents and teachers to rate the items and objectives in terms of their appropriateness for the students and the curriculum. The Rasch model item analysis rejected fifteen of the forty test items and a factor analysis of the test indicated that the rejected items were not measures of the first factor measured by the test. There was better than chance agreement between the item analysis and the ratings by teachers or superintendents. Very few items were accepted by the item analysis criterion and rejected by the rating criteria. It was concluded that the best evaluation of the items would include both criteria. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185111

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

WHAT DOES IT MEAN FOR ITEMS TO FIT A LATENT TRAIT PSYCHOMETRIC MODEL: A COMPARISON OF MODEL FITTING WITH CONTENT VALIDATION

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R.A. Garcia-Quintana

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Roan A. Garcia-Quintana
Program Specialist - Statistician
Office of Research
South Carolina Department of Education
(Presenter)

Joseph P. Ryan
Associate Professor
College of Education
University of South Carolina

Prepared for presentation at the Eastern Educational Research Association Conference, Kiawah Island, South Carolina, February 20-24, 1979

TM800 124

What Does It Mean For Items to Fit a Latent Trait Psychometric Model: A Comparison of Model Fiting With Content Validation

Introduction

During the past several years, thirty-six (36) states have set educational standards requiring students to exhibit minimum competence and many states use this information to determine whether a child should advance to the next grade and ultimately receive a high school diploma. In most situations, a test is selected or developed and students are required to attain a certain score before he or she is promoted to the next grade or awarded certification. The content validation of the tests used for this purpose is, of course, a very crucial procedure. Decisions based on tests which are not content valid are bound to be in error and can lead to serious damages to students.

In 1977, the South Carolina Legislation, under the School Finance Act, stated that "the State Department of Education shall develop or select and field test a competency based student assessment program in the basic skill areas of reading and mathematics utilizing criterion-referenced tests." This test was to be administered to students in Grades 3, 6, and 8. A committee of 26 educators, appointed by the State Superintendent, made the selection of the criterion-referenced tests (Thirteen members served on the Reading Subcommittee and thirteen served on the Mathematics Subcommittee). Both subcommittees consisted of members of the State Textbook Adoption Committee for the respective subject areas as well as parents, teachers, principals, superintendents and research coordinators.

Four tests were evaluated and both subcommittees decided the criterion-referenced test published by Science Research Associates (SRA) was the most appropriate for field testing. Although the decision made by each

subcommittee reflected a concensus of opinion about the objectives, there was still a need for superintendents and teachers to validate the test since objectives within the state vary from one school district to another.

In this study, only the SRA Sixth Grade Mathematics Test, Form X, will be examined. This test consists of 40 objectives, each measured by three items. The study examines the relationship between the fit of the test or objective to the Rasch latent trait model and the content validity of the objectives as identified by teachers and superintendents. This investigation explores both the nature of the unidimensional latent trait identified through latent trait analysis and also studies the feasibility of using the Rasch latent trait model in analyzing the content validity of tests in an applied educational setting.

Procedures

To determine the test's content validity with respect to the curriculum, all district superintendents in South Carolina as well as teachers whose students were involved in the pilot-testing were asked to judge each test. The objectives, as well as each of the three items measuring each objective, were judged.

The superintendents' evaluations were obtained through the use of the following questions:

(1) Does the objective represent the curriculum?

(2) Do you think the objective is essential?

(3) Is the item appropriate for the objective?

(4) Is the item appropriate for students?

The teachers' evaluations were obtained through the use of the following questions:

(1) Does the objective represent your instruction?

(2) Do you teach this objective?

(3) Is the item appropriate for the objective?

(4) Is the item appropriate for students?

For this analysis, the results from all of these questions will be examined.

The Rasch one parameter latent trait model for dichotomous responses will be used as a criteria for determining whether or not objectives fit an undimensional latent trait. The relationship between the objectives fit to the latent trait model and the degree of agreement among teachers and superintendents with respect to content validity is examined. The items on this test were examined for fit to the Rasch Model using the BICAL program of Wright and Mead.

Rather than examining individual items, a data set was created containing the number of items a particular subject answered correctly for each individual objective. Hence, for any one objective, a student may have a score of 0, 1, 2, 3, or left blank. Since this procedure was established for analyzing the student data, the percent agreement on item evaluation was averaged for the three items measuring each objective. Thus, an average percent agreement from the item evaluation is compared to the fit of that particular objective.

Sample

The subjects in this study are 700 children randomly selected from the sixth grade students in South Carolina to whom the SRA Mathematics Test, Form X, was administered in Spring, 1978. The superintendents' evaluations were based on

a sample size of 52 district superintendents. For this particular form of the test, sixty-six teachers responded to the item evaluation and 117 teachers responded to the evaluation of objectives. This discrepancy in the number of teacher respondents from item to objective evaluation is due to the fact that while two forms of the test were administered to each grade, both forms had the same objectives in common. Teachers were only asked to evaluate the items of the particular test they were to administer.

Analysis

The percent of agreement for the sample of superintendents and for the sample of teachers is based on the number of "yes" responses to each of the questions in the evaluation divided by the total number of responses. In most cases, this percent of agreement was very high (.90 or greater). The range of percent agreement varied somewhat for each question and for each type of respondent (i.e., superintendent or teacher).

Prior to examining the fit of each objective to the Rasch Model, the degree of agreement for each question on the teacher/superintendent evaluation was closely examined in order to establish a criterion point for rejection and acceptance (endorsement). Since the range of agreement for the different questions varied in the evaluation from one question to the other, the criterion for rejection and acceptance used to evaluate each objective differs from question to question. For both the teacher and superintendent evaluations, objectives were classified as "accepted" if the percent of agreement was greater than the criterion point and "rejected" if it was less than or equal to the criterion point.

Due to the small variation and high percent agreement among teachers and superintendents, some questions from the evaluation were not used for this

analysis. Both teachers and superintendents exhibited a very high percent agreement on the question, "Is the item appppropriate for the objective?". The percent of agreeement for both teacher and superintendents on this question for the objectives ranged from 95% to 100%. The superintendent question, "Do you think the objective is essential?", had an extremely high percent of agreement and thus was deleted from the analysis also.

Although the percent of agreement for the other evaluation questions was quite high for the majority of the objectives, at least some objectives exhibited a degree of agreement that could be labeled "rejected".

The Rasch analysis was performed on the objectives with scores of 0, 1, and 2 counted as "wrong" or "non-mastery", and scores of 3 counted as "correct" or "mastery". The objectives not fitting the model were compared to the objectives "rejected" by the teacher/superintendent evaluations.

Results

Tables 1-5 reveal the relationship between teacher or superintendent "acceptance" or "rejection" and the Rasch Model "acceptance" (or fit) and "rejection" (did not fit). In each case, over 70% of the objectives that were "accepted" or "rejected" by the teacher/superintendent evaluations were identically "accepted" or "rejected" by the Rasch Model. There were very few cases for which the Rasch Model "accepted" an objective which was "rejected" by the teacher/superintendent evaluation. Most of the inconsistencies between educators and the Rasch Model occurred when the Rasch Model "rejected" an objective which the teacher/superintendent evaluation had "accepted".

Since there seemed to be a "significant" number of objectives which the Rasch model "rejected" and the teachers and superintendents had "accepted", a

factor analysis was conducted in order to further understand the issues involved. The factor analysis revealed that the instrument was certainly not measuring an unidimensional trait. The first factor accounted for barely 25% of the variation. Perhaps the most interesting finding was that all items which did not load on the first factor were all "rejected" by the Rasch Model. Clearly this test was not homogeneous in the classical sense. This could account for the higher rejection rate of the Rasch Model as compared to the teacher evaluation.

Conclusions

The consistency of "rejected" or "accepted" objectives between the latent trait procedure and educators judgement was of moderate degree. For more than 70% of the objectives, teachers and superintendents evaluations of the objectives agreed with the fit of these objectives to the Rasch Model. Of peculiar interest is the fact that in very few instances did the Rasch Model accept an objective rejected by the teachers. Several objectives were rejected by the Rasch Model which teachers had found acceptable. It is important to remember, however, that there is no particular evidence to argue that the teacher/superintendent evaluations were unquestionably correct.

The Rasch Model proves to accept and reject at a rate which can be useful to educators in a setting where revisions of items or objectives are in order. The study suggests that educators' judgements and latent trait analysis might be most useful when used together. But when resources and time are scarce, the Rasch Model seems to perform most adequately.

References

South Carolina Validation Study of Criterion-Referenced Tests In Reading and Mathematics, Office of Research Report Series, South Carolina Department of Education, August, 1978.

Wright, B. D. and Mead, R. J. BICAL: Calibrating Items and Scales With the Rasch Model. Research Memorandum No. 23A, Statistics Laboratory, Department of Education, University of Chicago, 1978.

Table 1

Superintendent Question: Does the objective represent the curriculum? (Objectives with Agreements $\leq$ 80% were classified as rejected by Superintendents.)

Superintendent	Rasch: Rejected	Rasch: Accepted	
Accepted	(11) *2,3,11,16,21,26,30 32,35,39,40	(24) *1,4,5,6,7,8,9,10,12 13,14,15,17,18,19,20 22,23,24,27,28,29,31 36	35
Rejected	(4) *33,34,37,38	(1) *25	5
	15	25	40

70% consistent classification between Rasch model fit and Superintendent evaluation.

*Objective Numbers.

Table 2

Superintendent Question: Is the item appropriate for students? (Average 3-item-per-objective Agreements $\leq$ 75% were classified as rejected by Superintendents.)

Superintendent	Rasch: Rejected	Rasch: Accepted	
Accepted	(10) *2,3,11,16,21,30, 35,37,39,40	(24) *1,4,5,6,7,8,9,10,12 13,14,15,17,18,19, 20,22,23,24,27,28, 29,31,36	34
Rejected	(5) *26,32,33,34,38	(1) *25	6
	15	25	40

72.5% consistent classifications between Rasch Model fit and Superintendent evaluation.

*Objective Numbers.

Table 3

Teacher Question: Is the item appropriate for students? (Average 3-item-per-objective Agreements $\leq$ 70% were classified as rejected by teachers.)

Superintendent	Rasch: Rejected	Rasch: Accepted	
Accepted	(6) *2,3,11,16,21,30	(21) *1,4,5,6,7,8,9,10,12,13,15,17,18,19,20,22,23,24 27,28,29	27
Rejected	(9) *26,32,33,34,35,37 38,39,40	(4) *14,25,31,36	13
	15	25	40

75% consistent classification between Rasch Model fit and teacher evaluation.

*Objective Numbers.

Table 4

Teacher Question: Does the objective represent your instruction? (Objectives with Agreements $\leq$ 70% were classified as rejected by teachers.

		Rasch: Rejected	Rasch: Accepted	
Teacher	Accepted	(7) *2,3,11,16,21,30,40	(23) *1,4,5,6,7,8,9,10,12,13, 14,15,17,18,19,20,22,23, 24,27,28,29,31	30
	Rejected	(8) *26,32,33,34,35,37,38, 39	(2) *25,36	10
		15	25	40

77.5% consistent classification between Rasch Model fit and teacher evaluation.

*Objective Numbers.

Table 5

Teacher Question: Do you teach the objective?
(Objectives with Agreements $\leq$ 80% were classified as rejected by teachers.)

		Rasch	
	Rejected	Accepted	
Teacher Accepted	(11) *2,3,11,16,21,26, 30,32,33,39,40	(24) *1,4,5,6,7,8,9,10,12,13,14, 15,17,18,19,20,22,23,24, 27,28,29,31,36	35
Teacher Rejected	(4) *34,35,37,38	(1) *25	5
	15	25	40

72.5% consistent classification between Rasch Model fit and teacher evaluation.

*Objective Numbers.